Sources:

<https://nlp.johnsnowlabs.com/>

<https://nlp.johnsnowlabs.com/docs/en/pipelines>

<https://techvidvan.com/tutorials/ways-to-create-rdd-in-spark/>

<https://medium.com/the-data-experience/building-a-data-pipeline-from-scratch-32b712cfb1db>

<https://dzone.com/articles/calculating-tf-idf-with-apache-spark>

<https://mapr.com/blog/similar-document-search-using-apache-spark-with-tf-idf/>

Problem 1:

<https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/4987911583861567/1369723412274537/3511878718732474/latest.html>

Problem 2:

<https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/4987911583861567/2319919622347593/3511878718732474/latest.html>